

Estatística Descritiva

O que é Estatística

- A Estatística originou-se com a coleta e construção de tabelas de dados para o governo.
- A situação evoluiu e esta coleta de dados representa somente um dos aspectos da Estatística.
- No século XIX, o desenvolvimento do cálculo de probabilidade e outras metodologias matemáticas, tais como a técnica de Mínimos Quadrados, foram fundamentais para o desenvolvimento da Estatística

O que é Estatística

- Somente no século XX a Estatística desenvolve-se como uma área específica do conhecimento a partir do desenvolvimento da Inferência Estatística; uma metodologia baseada em probabilidade que tem ampla aplicação nas ciências experimentais.
- A Estatística hoje consiste num metodologia científica para obtenção, organização e análise de dados, oriundos das mais variadas áreas das ciência experimentais, cujo objetivo principal é auxiliar a tomada de decisões em situações de incerteza.

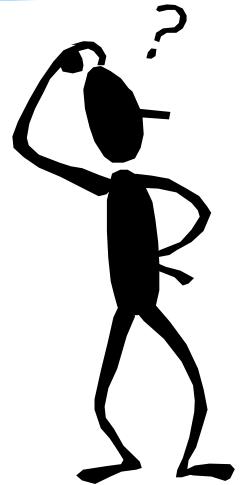
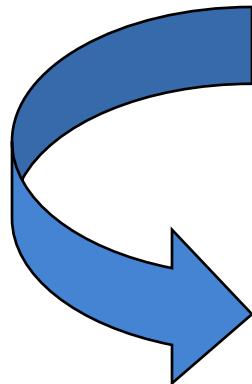
Estatística Descritiva

Etapa inicial da análise utilizada para descrever, organizar e resumir os dados coletados.

A disponibilidade de uma grande quantidade de dados e de métodos computacionais muito eficientes revigorou esta área da Estatística.

Estatística Descritiva

O que fazer com as observações
que coletamos?



Primeira Etapa:

Resumo dos dados = Estatística descritiva

Variável:

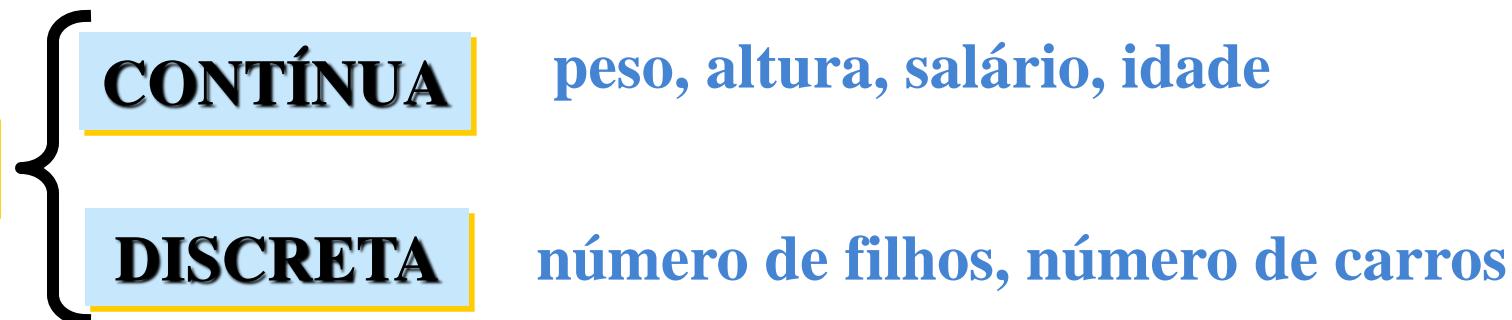
Qualquer característica associada a uma população.

Classificação das variáveis:

QUALITATIVA



QUANTITATIVA



Variáveis Quantitativas

MEDIDAS DE POSIÇÃO:

Mínimo, Máximo, Moda, Média, Mediana, Percentis

MEDIDAS DE DISPERSÃO:

Amplitude, Intervalo-Interquartil, Variância, Desvio Padrão, Coeficiente de Variação.

Medidas de Posição

- **Máximo (max)**: a maior observação
- **Mínimo (min)**: a menor observação
- **Moda (mo)**: é o valor (ou atributo) que ocorre com maior freqüência.

Dados: 4, 5, 4, 6, 5, 8, 4

$$\text{max} = 8$$

$$\text{min} = 4$$

$$\text{mo} = 4$$

•Média:

$$\bar{x} = \frac{x_1 + x_2 + x_3 + \dots + x_n}{n} = \frac{\sum_{i=1}^n x_i}{n}$$

Dados: 2, 5, 3, 7, 8

$$\bar{x} = \frac{2+5+3+7+8}{5} = 5$$

- **Mediana:**

A mediana é o valor da variável que ocupa a posição central de um conjunto de n dados ordenados.

Posição da mediana: $\frac{n+1}{2}$

Exemplos:

Dados: 2, 6, 3, 7, 8

$\Rightarrow n = 5$ (ímpar)

Dados ordenados: 2 3 6 7 8 $\Rightarrow \frac{5+1}{2} = 3 \Rightarrow Md=6$

Posição da Mediana ↑

Dados: 4, 8, 2, 1, 9, 6

$\Rightarrow n = 6$ (par)

Dados ordenados: 1 2 4 6 8 9 $\Rightarrow \frac{6+1}{2} = 3,5$

Md

$$Md = (4 + 6) / 2 = 5$$

Percentis:

*O percentil de ordem $p \times 100$ ($0 < p < 1$), em um conjunto de dados de tamanho n , é o valor da variável que ocupa a posição $p \times (n + 1)$ do conjunto de dados ordenados.

Casos particulares:

percentil 50 = mediana ou segundo quartil (Md)

percentil 25 = primeiro quartil (Q_1)

percentil 75 = terceiro quartil (Q_3)

percentil 10 = primeiro decil

Dados: 1,9 2,0 2,1 2,5 3,0 3,1 3,3 3,7 6,1 7,7 $\Rightarrow n=10$

Posição de Md : $0,5(n+1) = 0,5 \times 11 = 5,5 \Rightarrow Md = (3 + 3,1)/2 = 3,05$

Posição de $Q1$: $0,25(11) = 2,75 \Rightarrow Q_1 = (2 + 2,1)/2 = 2,05$

Posição de $Q3$: $0,75(11) = 8,25 \Rightarrow Q_3 = (3,7 + 6,1)/2 = 4,9$

$$Md = 3,05$$

$$Q_1 = 2,05$$

$$Q_3 = 4,9$$

Dados: 0,9 1,0 1,7 2,9 3,1 5,3 5,5 12,2 12,9 14,0 33,6

$\Rightarrow n=11$

$$Md = 5,3$$

$$Q1 = 1,7$$

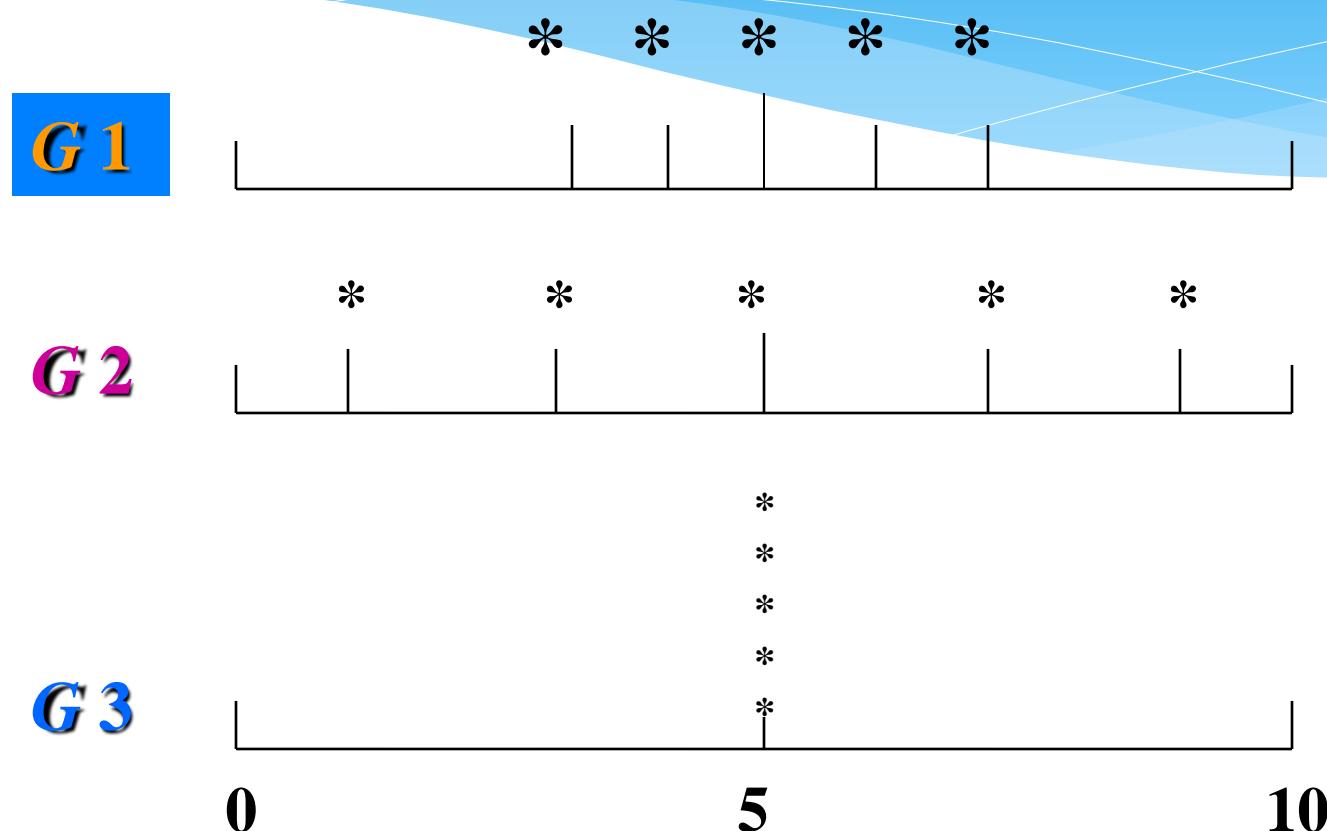
$$Q3 = 12,9$$

Exemplo 2: Considere as notas de um teste de 3 grupos de alunos

Grupo 1: 3,4,5,6,7

Grupo 2: 1, 3, 5, 7, 9

Grupo 3: 5,5,5,5,5



Temos: $\bar{x}_1 = \bar{x}_2 = \bar{x}_3 = 5$ e $md_1 = md_2 = md_3 = 5$

Medidas de Dispersão

Finalidade: encontrar um valor que resuma a variabilidade de um conjunto de dados

- **Amplitude (A):**

$$A = \text{máx} - \text{min}$$

Para os grupos anteriores, temos:

$$\text{Grupo 1, } A = 4$$

$$\text{Grupo 2, } A = 8$$

$$\text{Grupo 3, } A = 0$$

•Intervalo-Interquartil:

É a diferença entre o terceiro quartil e o primeiro quartil, ou seja, $Q3 - Q1$.

Dados: 1,9 2,0 2,1 2,5 3,0 3,1 3,3 3,7 6,1 7,7

$$Q1 = 2,05 \quad \text{e} \quad Q3 = 4,9$$

$$Q3 - Q1 = 4,9 - 2,05 = 2,85$$

•Variância:

$$\text{Variância} = s^2 = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{n-1} = \sum_{i=1}^n \frac{(x_i - \bar{x})^2}{n-1}$$

•Desvio padrão:

$$\text{Desvio Padrão} = s = \sqrt{\text{Variância}}$$

Cálculo para os grupos:

$$\textbf{G1: } s^2 = \frac{(3-5)^2 + (4-5)^2 + (5-5)^2 + (6-5)^2 + (7-5)^2}{4}$$

$$\Rightarrow s^2 = 10/4 = 2,5 \Rightarrow s = 1,58$$

$$\textbf{G2: } s^2 = 10 \Rightarrow s = 3,16$$

$$\textbf{G3: } s^2 = 0 \Rightarrow s = 0$$

Fórmula alternativa:

$$S^2 = \frac{\sum_{i=1}^n X_i^2 - n\bar{X}^2}{(n - 1)}$$

Em G1: $\Sigma X_i^2 = 9 + 16 + 25 + 36 + 49 = 135$

$$\Rightarrow S^2 = \frac{135 - 5 \times (5)^2}{4} = 2,5$$

• Coeficiente de Variação (CV)

- é uma medida de dispersão relativa
- elimina o efeito da magnitude dos dados
- exprime a variabilidade em relação à média

$$CV = \frac{s}{\bar{x}} \times 100\%$$

Exemplo 3:

Altura e peso de alunos

	Média	Desvio Padrão	Coef. de Variação
Altura	1,143m	0,063m	5,5%
Peso	50 kg	6kg	12%

Conclusão: Os alunos são, aproximadamente, duas vezes mais dispersos quanto ao peso do que quanto à altura.

Exemplo 4:

Altura (em *cm*) de uma amostra de recém-nascidos e de uma amostra de adolescentes

	Média	Desvio padrão	Coef. de variação
Recém-nascidos	50	6	12%
Adolescentes	160	16	10%

Conclusão: Em relação às médias, as alturas dos adolescentes e dos recém-nascidos apresentam variabilidade quase iguais.